Diversity, Conflict and Agglomeration in African Cities

Andre Gray

February 22, 2024

Motivation

- Africa is a diverse continent in terms of ethnicities, languages, religions
- past work suggests diversity in Africa has costs
  - $\uparrow$ diversity at state level $\rightarrow$ lower economic growth, more conflict (Arbatli et al., 2020; Alesina and Ferrara, 2005; Mueller et al., 2022; Esteban et al., 2012)
- Africa is also rapidly urbanizing
  - in US/Europe we think of large cosmopolitan cities as growth centers (returns to density)
  - is the same true in Africa?

Research Question

- how does diversity help or hinder development in African cities?
  - **positive returns:** love of variety, ethnicity-specific knowledge and ideas (Montalvo and Reynal-Querol, 2021; Mueller et al., 2022)
  - **negative returns:** ethnic diversity works as a congestion force, dampens agglomeration benefits of density through conflict or lower productivity
  - **size effect?** maybe large cities able to manage ethnic diversity through firm-ethnicity specialization, residential sorting (Glaeser et al., 1995)
- the answer to this Q has implications for returns to urbanization in Africa

This project

- Empirical approach
  - compare the development of cities with different ex ante levels of fractionalization
  - leverage pull (labor demand shocks) and push (climate disasters) factors for exogenous variation in exposure to fractionalization
- Results
  - cities located in relatively more diverse regions are smaller, have lower light density, and more conflict.
- Model
  - motivate a spatial equilibrium model with upward sloping labor supply where cities compete for workers from discrete ethnic groups (Monte et al., 2018; Diamond, 2016)
  - cities receive agglomeration benefits from density, but trade-off congestion costs of ethnic violence as a function of population diversity

Example 1 Example 2

Endogenous Amenities with Many Groups: Set-up

Rosen-Roback with heterogeneous ethnic groups and conflict:

- S locations indexed by i
- workers from J discrete groups, where g_{ji} is number of workers in group j in region i
- many firms with free entry: $y_i = A_i L_i$
- Productivity of a firm d at location i has a conflict cost C_i : $A_{di} = \bar{A}_i L_{di}^{\alpha} C_i(g_{1i}, g_{2i} ... g_{Ji})^{-\gamma}$

Endogenous Amenities with Many Groups: Labor Demand

- Because conflict is defined at the city level, each individual firm takes this cost as given and simply chooses a number of L workers such that

$$W_{di} = (\alpha + 1)\bar{A}_i L_{di}^{\alpha} C_i(g_{1i}, g_{2i}...g_{Ji})^{-\gamma} \quad (1)$$

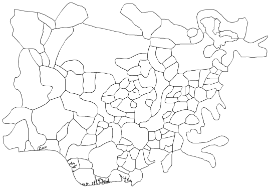
Adding up across firms in location i we have total labor demand in city i:

$$ln(W_i) = ln(\alpha + 1) + ln(\bar{A}_i) + \alpha ln(L_i) - \gamma ln(C_i(g_{1i}, g_{2i}...g_{Ji})) \quad (2)$$

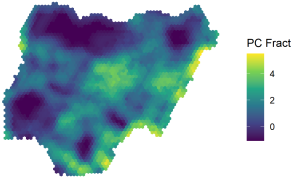
Labor Supply | Production

Empirics

Ethnic Fractionalization



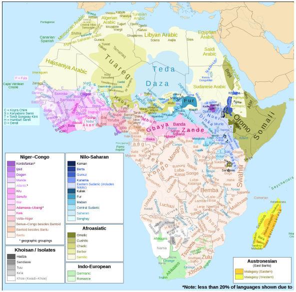
(a) Nigeria (Murdock Map, 1967)



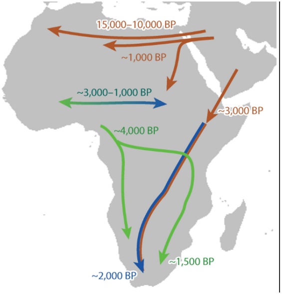
(b) PC Fractionalization

Data | Correlations | Diversity Measure | Validate Diversity Measure | Population data

Origins of Ethnic Diversity



(a) Language Families



(b) (Schlebusch and Jakobsson, 2018)

- ancient migrations (Bantu Expansion, Eurasian backflow, spread of Islam)
- function of geography (land suitability, natural borders) – isolation creates new splintered groups (Michalopoulos, 2012)

Pull and Push Shocks

- pull factors *(historic fractionalization)*
  - regions experience a shock to productivity independent of underlying ethnic diversity
  - **railroads, portage sites**, natural harbors, mining discovery
- push factors *(contemporary fractionalization)*
  - exogenous shocks push migrants into cities
  - varying impact on city diversity
  - **droughts**, floods, commodities

Model to Empirics

- from labor demand we want a proxy for object C we can estimate

$$ln(W_i) = ln(\alpha + 1) + ln(\bar{A}_i) + \alpha ln(L_i) - \gamma ln(C_i(g_{1i}, g_{2i}...g_{Ni})) \quad (3)$$

- use a region's underlying potential diversity Div_i based on historical ethnic groups, interacted with labor demand shock

$$ln(W_i) = ln(\alpha + 1) + ln(\bar{A}_i) + \alpha ln(L_i) - \gamma Div_i * ln(L_i) + \omega Div_i \quad (4)$$

Identification 1: Railroad Towns and Least-Cost Path

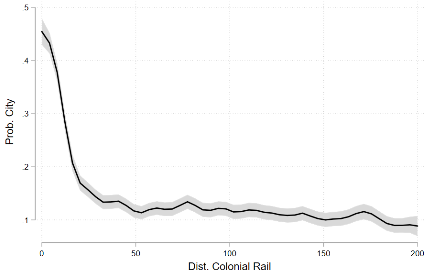
Identification 1: Railroad Towns and Least-Cost Path

- colonial era railroads built to connect ports to natural resources of interest
- railroads often followed least cost-path to copper mine or other resource (Jedwab et al., 2017)
- regions along railroad path experience a productivity shock by new rail access, workers migrate in from hinterland
- cities emerged along the railroad path due to access, **plausibly independent of location fundamentals and ethnic distribution of hinterland**

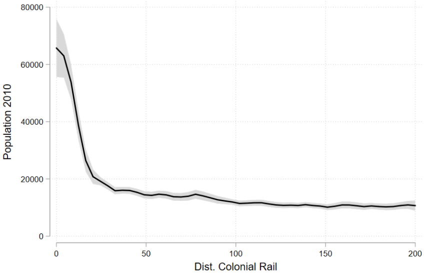
Example Rail Map

Identification 1: Railroad Towns and Least-Cost Path

Figure: Local Polynomial of City Formation by Dist to Rail (km)



(a) Prob. of City Location



(b) City Pop. 2010

Identification 1: First Stage

First-stage:

Location i in state s near rail line r with diversity Div_i

$$L_i = \alpha + \beta_1 Dist_i + \upsilon_r + \omega_s + \epsilon_i \quad (5)$$

$$L_i * Div_i = \alpha + \beta_1 Dist_i + \beta_2 Dist_i * Div_i + \upsilon_r + \omega_s + \epsilon_i \quad (6)$$

Second stage:

$$y_i = \alpha + \beta_2 \hat{L_i} + \beta_3 Div_i + \beta_4 L_i \widehat{*} Div_i + X_i + \epsilon_i \quad (7)$$

υ_r, ω_s control for railway and state fixed effects respectively.

Corr with Diversity 1 | Corr with Diversity 2 | First Stage

Identification 1: Second Stage

Table: Rail IV - Light Density 2010s

	Lights	Lights	Lights	Lights
City*PC Fract	-0.133	-0.239	-0.243	-0.236
	[0.064]**	[0.080]***	[0.099]**	[0.105]**
PC Fract	0.035	0.050	0.074	0.059
	[0.015]**	[0.017]***	[0.029]**	[0.029]**
Rail FE	N	Y	N	Y
Dist to Rail	<300km	<300km	<100km	<100km
F-stat	358	181	222	197
Mean Dep. Var	-0.030	-0.030	0.006	0.006
Observations	40,257	40,257	17,268	17,268

Notes: estimates from equation $y_i = \alpha + \beta_2 \hat{L_i} + \beta_3 Div_i + \beta_4 L_i * \hat{} Div_i + X_i + \epsilon_i$

Conflict Results Durables

Identification 2: Portage Sites

Portage Sites

- portage sites – towns often built near point at which a river becomes non-navigable (Bleakley and Lin, 2012)
- need infrastructure at these points to switch goods from ship to land or riverboat
- applied to Africa:
  - use data on elevation, ruggedness and river depth across Africa (Nunn and Puga, 2012)
  - identify all potential sites as an instrument for city locations
  - need these sites to be orthogonal to distribution of historical ethnic settlements

Example Portage Map

Predicting City Location with Portage Sites

Construct measure of portage propensity R_i as a standardized interaction of distance to river and ruggedness.
First stage:

$$L_i = \alpha + \beta_1 R_i + X_i + \upsilon_s + \epsilon_i \tag{8}$$

$$L_i * Div_i = \alpha + \beta_1 R_i * Div_i + X_i + \upsilon_s + \epsilon_i \tag{9}$$

Second stage:

$$y_i = \alpha + \beta_2 \hat{L_i} + \beta_3 Div_i + \beta_4 L_i \hat{*} Div_i + X_i + \epsilon_i \tag{10}$$

Corr with Diversity | First Stage | Validation

Identification 2: Second Stage

Table: Portage IV - Light Density 2010s

	Lights	Lights	Lights	Lights
City*PC Fract	-0.017	-0.334	-0.066	-0.211
	[0.178]	[0.166]**	[0.367]	[0.260]
PC Fract	0.003	0.058	0.012	0.027
	[0.037]	[0.034]*	[0.074]	[0.052]
River FE	N	Y	N	Y
Dist to River	<100km	<100km	<50km	<50km
F-stat	42	62	20	38
Mean Dep. Var	-0.002	-0.002	0.037	0.037
Observations	36,747	36,747	22,861	22,861

Notes: Controls include land suitability, malaria suitability. All regressions include country fixed effects. Fractionalization measures are standardized, and defined using a 50km buffer from the grid centroid. Light density measures are also standardized after averaging across years 2000-2009 and 2010-2013.

Conflict Outcomes Durables

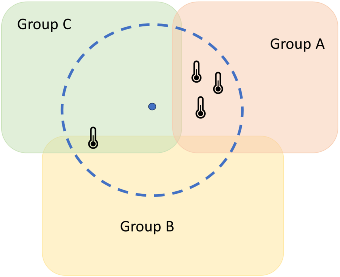
Push Events and Modern Diversity

Shocks to Population and Diversity

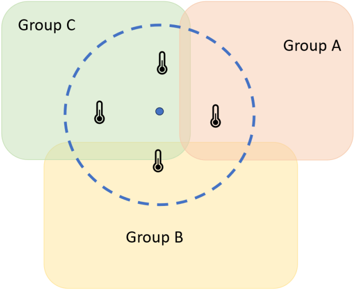
- use exposure to drought shocks to instrument for population and diversity change

$$ln(W_i) = ln(\alpha + 1) + ln(\bar{A}_i) + \alpha ln(L_i) - \gamma ln(C_i(g_{1i}, g_{2i}...g_{Ni})) \quad (11)$$

Relative Drought Exposure



(a) Droughts concentrated in one group



(b) Droughts uniformly distributed

- instrument for population increase by total drought intensity
- instrument for diversity by relative distribution of drought across groups in catchment area

Conclusion

- this papers explores how regional diversity mitigates the impact of exogenous labor demand shocks on city growth
- portage sites and colonial railroads are used as plausibly exogenous shocks to regional demand unrelated to historical diversity
- I find that cities that are placed in more diverse areas see more conflict and have lower light density

References I

Alesina, A. and Ferrara, E. L. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, 43(3):762–800.

Alesina, A. and Zhuravskaya, E. (2011). Segregation and the quality of government in a cross section of countries. *American Economic Review*, 101(5):1872–1911.

Arbatli, C. E., Ashraf, Q. H., Galor, O., and Klemp, M. (2020). Diversity and Conflict. *Econometrica*, 88(2):727–797.

Beck, J. and Sieber, A. (2010). Is the spatial distribution of mankind's most basic economic traits determined by climate and soil alone? *PloS one*, 5(5):e10416.

Bleakley, H. and Lin, J. (2012). Portage and Path Dependence *. *The Quarterly Journal of Economics*, 127(2):587–644.

Bryan, G. and Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy*, 127(5):2229–2268.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. *American Economic Review*, 106(3):479–524.

Esteban, J., Mayoral, L., and Ray, D. (2012). Ethnicity and Conflict: An Empirical Study. *American Economic Review*, 102(4):1310–1342.

Glaeser, E. L., Scheinkman, J., and Shleifer, A. (1995). Economic growth in a cross-section of cities. *Journal of monetary economics*, 36(1):117–143.

Jedwab, R., Kerby, E., and Moradi, A. (2017). History, Path Dependence and Development: Evidence from Colonial Railways, Settlers and Cities in Kenya. *The Economic Journal*, 127(603):1467–1494.

References II

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S. E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. *The American journal of tropical medicine and hygiene*, 70(5):486–498.

Li, X., Zhou, Y., Zhao, M., and Zhao, X. (2020). A harmonized global nighttime light dataset 1992–2018. *Scientific data*, 7(1):168.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *American Economic Review*, 102(4):1508–1539.

Montalvo, J. G. and Reynal-Querol, M. (2021). Ethnic Diversity and Growth: Revisiting the Evidence. *The Review of Economics and Statistics*, 103(3):521–532.

Monte, F., Redding, S. J., and Rossi-Hansberg, E. (2018). Commuting, migration, and local employment elasticities. *American Economic Review*, 108(12):3855–90.

Mueller, H., Rohner, D., and SchÃ¶nholzer, D. (2022). Ethnic Violence Across Space. *The Economic Journal*, 132(642):709–740.

Murdock, G. P. (1967). Ethnographic atlas: a summary. *Ethnology*, 6(2):109–236.

Nunn, N. and Puga, D. (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics*, 94(1):20–36.

Schlebusch, C. M. and Jakobsson, M. (2018). Tales of human migration, admixture, and selection in africa. *Annual Review of Genomics and Human Genetics*, 19:405–428.

Heterogeneity by City Size

Diversity and City Size

Figure: Diversity Association with Nighttime Lights by City Size

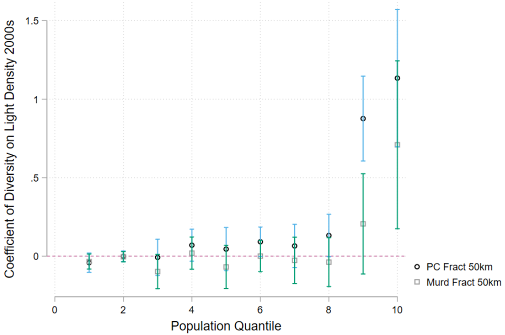


Figure: Grids in Africapolis Cities Dataset

Diversity and City Size

- potential research questions on nonlinear relationship between diversity, city size and growth:
  - relationship between residential segregation and industry/job segregation?
  - does firm-level discrimination vary by firm size, market structure?
  - do production characteristics (ex. labor intensity), affect segregation?
  - does segregation vary over city life cycle (ie assimilation)

Industry-level Segregation

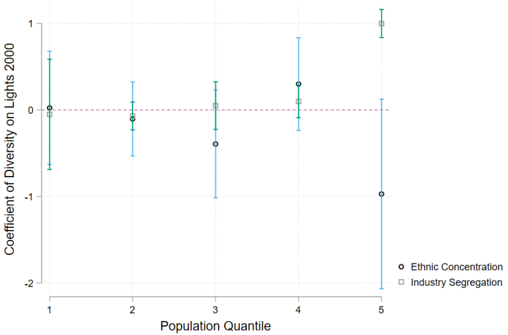
- are larger cities more segregated, and does segregation change the gains/losses of diversity?
- use census data at county level for available countries
- follow Alesina and Zhuravskaya (2011) to measure segregation of ethnic groups $j \in J$ across industries $i \in I$:

$$Seg_c = \frac{1}{J-1} \sum_{j=1}^{J} \sum_{i=1}^{I} \frac{N_j}{N_c} \frac{(\pi_{ij} - \pi_j)^2}{\pi_j} \tag{12}$$

Where π_j is the fraction of group j in county c, and π_{ji} is the fraction of group j in industry i of county c.

Diversity and City Size

Figure: Census Diversity and Light Intensity by Population



Note: Regression of $Lights_c = \alpha + \beta_1 X_c + \epsilon_c$, run by population quintile. The variable X is either Ethnic HHI or county level segregation respectively.

Diversity and City Size

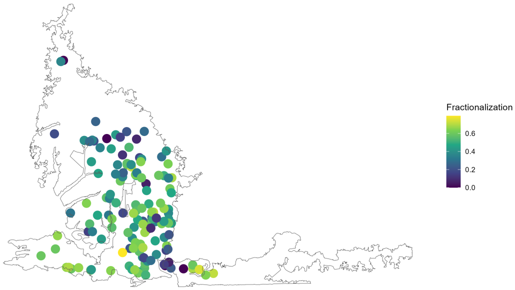
Table: Census Diversity and Light Intensity

	Lights 2000	Lights 2000	Lights 2000
Industry Concentration (HHI)	-2.737		
	[0.291]***		
Ethnic Concentration (HHI)		-0.368	
		[0.249]	
Industry-level Segregation			0.592
			[0.145]***
Population	0.682	0.654	0.426
	[0.180]***	[0.183]***	[0.130]***
Urban	-0.565	0.314	0.230
	[0.113]***	[0.061]***	[0.046]***
Observations	2,384	2,384	2,253

Note: This table shows coefficients from a regression of $Lights_c = \alpha + \beta_1 X_c + \epsilon_c$, controlling for current population, census year, dummy for urban region, and fixed effects for state and country. Data is at county level (administrative level 2). The variable X is the Industry HHI, Ethnic HHI and county level segregation respectively.

Residential Segregation

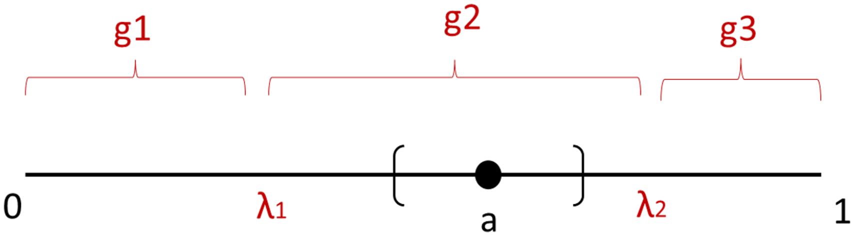
Figure: DHS Clusters within Lagos



Within-City Entropy Index

- where π_{jw} is the fraction of neighborhood w from group j
- entropy index : $E_w = \sum_1^J \pi_{jw} ln(\pi_{jw})$
- city c entropy is : $\frac{E_c - \bar{E_w}}{E_c}$

A Model of Diversity and City Labor Supply



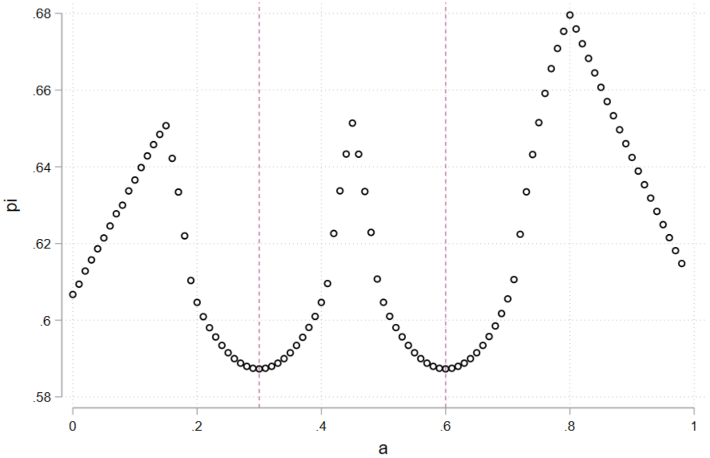
- city located at a draws workers from nearby regions
  - city produces and creates profit according to

$$AL^{\alpha} - pL - C(w_{g1}, w_{g2}, w_{g3}) \quad (13)$$

- workers come from ethnicities $g1, g2, g3$
  - proportions given by λ_1, λ_2
  - if worker in g_1 at location x moves to city, gets $p - t(|a - x|) + \beta(w_{g1}/L)$

return

A Model of Diversity and City Labor Supply



Example: Workers internalize no congestion cost ($\beta = 0$, city faces conflict cost
$F = \sum_1^3 \frac{w_i}{L}(1 - \frac{w_i}{L})$)

return

Endogenous Amenities with Many Groups: Labor Supply

- Workers from ethnic group j are born in region o and decide where to reside i, where they receive the equilibrium wage w_i.
- location-specific amenity negatively related to total population $L_i^{-\beta}$, and positively related to their relative group's share in the local labor force $(\frac{g_{ji}}{L_i})^{\upsilon}$
- individual worker t also receive an idiosyncratic preference shock for each region ϕ_{it} that is distributed Frechet $F(z) = e^{-z^{-\theta}}$
- Moving from origin o to i incurs a migration cost τ_{oi}
- Given the above, worker t from group j and birthplace o receives the following total utility if they move to i:

$$U_{jiot} = \frac{w_i}{P_i} L_i^{-\beta} (\frac{g_{ji}}{L_i})^{\upsilon} \tau_{oi} \phi_{it} \tag{14}$$

Where P_i is the local price index.

return

Endogenous Amenities with Many Groups: Production

- Following Bryan and Morten (2019), I assume that a representative firm maximizes the economy-wide production Y that aggregates the regional varieties y_i according to :

$$Y = (\sum_{i=1}^{S} y_d^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \tag{15}$$

The individual prices for regional products p_d are pinned down by the representative firm maximizing the production of Y subject to costs $\sum_d p_d y_d$. This gives us $p_d = (\frac{Y}{y_d})^{\frac{1}{\sigma}}$. Individual workers consume the final good Y, and it enters linearly into utility. We take this price as the numeraire.

return

Data

Outcomes:

- regional light density: DMSP and VIIRS harmonized (Li et al., 2020)
- probability of violent conflict: UCDP conflict events data

Ethnic Diversity:

- Murdock ethnic boundaries (Murdock, 1967),Soviet Narodov Mira atlas (GREG), Ethnologue
- IPUMS DHS (all years) and Census
- Afrobarometer

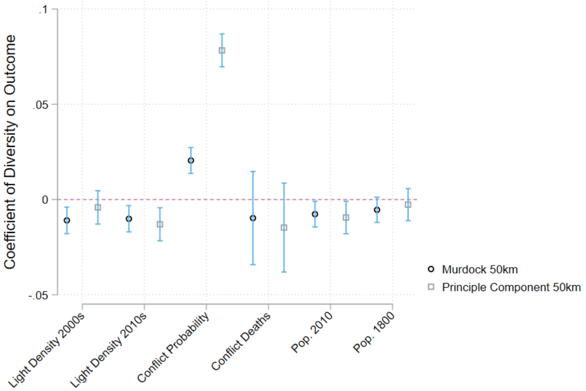
Settlements:

- African city settlement data 1950-2015: OECD/SWAC (2020), Africapolis (database), www.africapolis.org
- Geographic: malaria ecology (Kiszewski et al., 2004); suitability (Beck and Sieber, 2010), river network (HydroSHEDs), ruggedness (Nunn and Puga, 2012), mineral deposits
- Worldpop, History Database of the Global Environment (HYDE)

return

Development and Diversity

Figure: Association of Diversity with Lights and Conflict



Notes: These coefficients are estimated from the regression $y = \alpha + \beta D + X + \upsilon_s$, where D is a standardized measure of diversity either Murdock fractionalization or the principle component.

return

Diversity Measures

- fractionalization index $F_i = \sum_1^J \frac{g_{ji}}{L_i}(1 - \frac{g_{ji}}{L_i})$
- for historical diversity, we substitute population for land area via ethnic maps:
  - 50m radius around grid
  - set proportions as fraction of circle occupied by ethnic group j

return

Validate Diversity Measure

Table: Relationship of Diversity Measures to Census Diversity

	Murd Fract	PC Fract	Murd Count	Fract Lang	Fract GREG
Census Ethnic Concentration	-0.691 [0.080]***	-0.394 [0.075]***	-0.491 [0.088]***	-0.104 [0.018]***	-0.092 [0.016]***
Observations	2,384	2,384	2,384	2,384	2,384

Notes: The census sample includes %10 samples from Benin (1979,1992,2002,2013), Ethiopia (1994), Ghana (2000,2010), Guinea (2014), Malawi (2008), Mali (2009), Mauritius (2000,2011), Morocco (2014), Senegal (2013), Sierra Leone (2004), Togo (2010), Uganda (2002), Zambia (2000, 2010). Ethnic concentration for county i is calculated as $\sum_{j=1}^{J}(\frac{g_j}{N_i})^2$ where g_j is the number of people from ethnic group j, and N_i is the total sampled population of the county. Higher levels of ethnic concentration imply less diversity. Regressions include country fixed effects.

return

Population Data: Africapolis 1950-2000

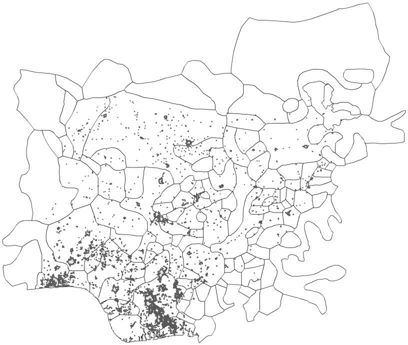


Figure: Towns/Settlements > 10k, Africapolis

return

Predicting Ethnic Diversity

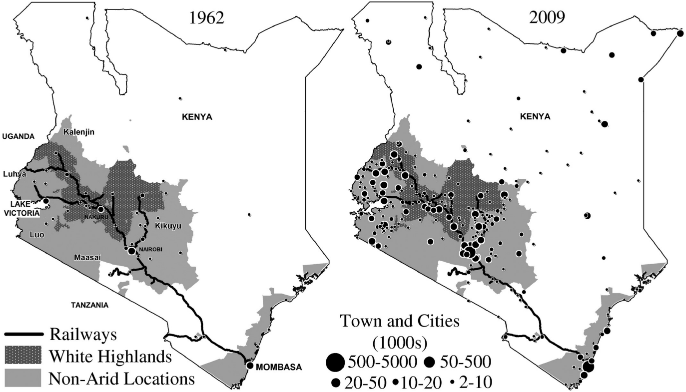
Table: Geography and Ethnic Diversity

	PC Fract	Murd Fract	Murd Count	Greg Count	Lang.
Dist River	-0.229	-0.157	-0.175	-0.426	0.080
	[0.005]***	[0.006]***	[0.006]***	[0.014]***	[0.002]***
Dist Coast	0.079	0.001	0.020	0.234	-0.029
	[0.005]***	[0.006]	[0.006]***	[0.016]***	[0.002]***
Malaria Suit.	0.135	0.083	0.121	0.242	-0.036
	[0.005]***	[0.005]***	[0.006]***	[0.014]***	[0.002]***
Pastoral Suit	-0.055	-0.021	-0.044	-0.022	-0.013
	[0.004]***	[0.004]***	[0.005]***	[0.012]*	[0.002]***
Agricultural Suit	0.186	0.175	0.219	0.247	-0.069
	[0.006]***	[0.006]***	[0.007]***	[0.018]***	[0.002]***
Past*Agr Suit	0.027	0.033	0.019	0.136	-0.016
	[0.004]***	[0.004]***	[0.004]***	[0.011]***	[0.001]***
Elevation	-0.082	-0.061	-0.056	-0.266	0.021
	[0.005]***	[0.005]***	[0.006]***	[0.012]***	[0.002]***
Ruggedness	0.086	0.034	0.057	0.266	-0.010
	[0.004]***	[0.004]***	[0.004]***	[0.011]***	[0.001]***
Mean Dep.	-0.018	-0.013	2.056	2.789	0.704
Observations	81,073	88,714	88,714	85,412	88,714

Notes: The fractionalization measures are standardized. The regressions include country fixed effects.

return

Identification 1: Railroad Towns and Least-Cost Path



Example: British rail connecting Uganda's Lake Victoria to coast for geopolitical reasons passed through Kenya incidentally (Jedwab et al., 2017)

return

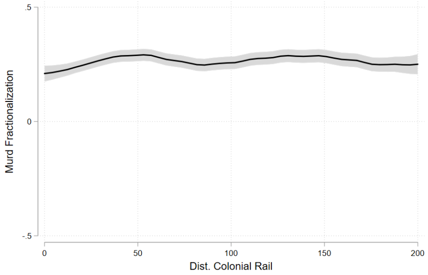
Identification 1: Railroad Towns and Least-Cost Path

Figure: Colonial rail lines (Jedwab et al., 2017)

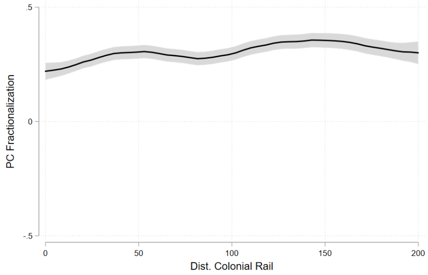
return

Identification 1: Railroad Towns and Least-Cost Path

Figure: Local Polynomial of Ethnic Fractionalization by Dist to Rail (km)



(a) Murdock Fractionalization



(b) PC Fractionalization

return

Identification 1: Railroad Towns and Least-Cost Path

Table: Fractionalization and Dist to Rail

	PC Fract	PC Fract	PC Fract	PC Fract
Dist to Rail <60km	-0.120	-0.089	-0.027	-0.012
	[0.011]***	[0.010]***	[0.014]*	[0.012]
Dist to Rail	<300km	<300km	<100km	<100km
Rail FE	N	Y	N	Y
Observations	40,008	40,008	17,173	17,173

Controls include land suitability, malaria suitability, ruggedness. Country and rail fixed effects.

return

Identification 1: First Stage

Table: Rail IV - Predict City

	Prob. City	Prob. City	Prob. City	Prob. City
Dist to Rail < 60km	0.096	0.078	0.075	0.069
	[0.004]***	[0.004]***	[0.006]***	[0.005]***
Dist to Rail	<300km	<300km	<100km	<100km
Rail FE	N	Y	N	Y
Mean Dep.	0.136	0.136	0.176	0.176
Observations	41,436	41,436	17,763	17,763

Notes: Controls include land suitability, malaria suitability, ruggedness. All regressions include country and rail fixed effects. Fractionalization measures are standardized, and defined using a 50km buffer from the grid centroid. The "Dist" row describes the sample cutoff of distance to nearest colonial rail for that particular regression.

return

Identification 1: Railroad Towns and Least-Cost Path

Table: Rail IV - Prob. Conflict

	P(conflict)	P(conflict)	P(conflict)	P(conflict)
City*PC Fract	0.003	0.011	0.019	0.030
	[0.005]	[0.006]*	[0.006]***	[0.006]***
PC Fract	0.005	0.000	-0.000	-0.006
	[0.001]***	[0.001]	[0.002]	[0.002]***
Rail FE	N	Y	N	Y
Dist to Rail	<300km	<300km	<100km	<100km
F-stat	358	181	222	197
Mean Dep. Var	0.014	0.014	0.012	0.012
Observations	40,257	40,257	17,268	17,268

Notes: Controls include land suitability, malaria suitability, ruggedness. All regressions include country and rail fixed effects. Fractionalization measures are standardized, and defined using a 50km buffer from the grid centroid. Prob. conflict is defined as the proportion of years in which the grid experienced a conflict across 1975-2021.

return

Identification 1: Railroad Towns and Least-Cost Path

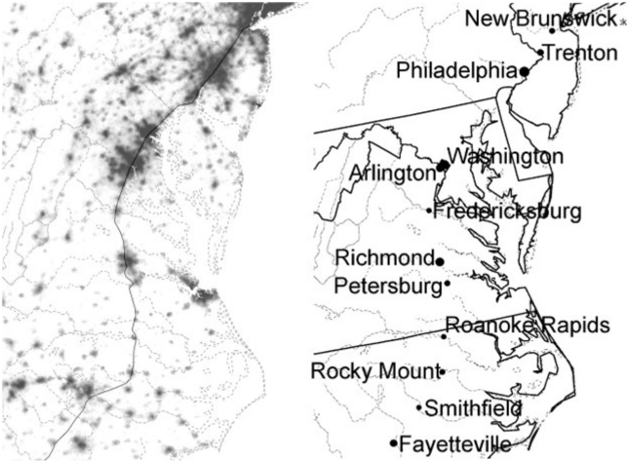
Table: Rail IV - DHS Durables

	Durables	Durables	Durables	Durables
City*PC Fract	-0.264	-0.098	-0.219	-0.204
	[0.010]***	[0.014]***	[0.016]***	[0.018]***
PC Fract	0.117	0.018	0.097	0.103
	[0.007]***	[0.009]*	[0.012]***	[0.014]***
Rail FE	N	Y	N	Y
Dist to Rail	<300km	<300km	<100km	<100km
F-stat	15166	5092	7874	5216
Mean Dep. Var	-0.001	-0.001	0.151	0.151
Observations	590,974	590,974	378,117	378,117

Notes: Controls include land suitability, malaria suitability, ruggedness. All regressions include country and rail fixed effects. Fractionalization measures are standardized, and defined using a 50km buffer from the grid centroid. Durables is a principle component analysis of assets reported in DHS samples.

return

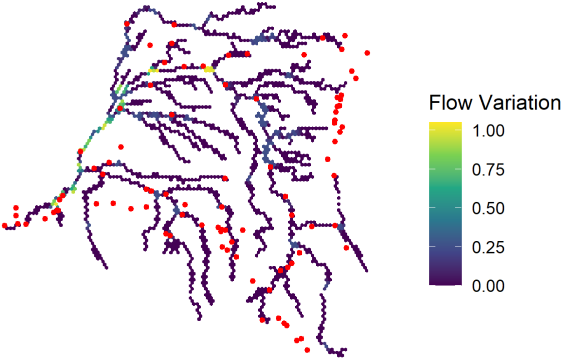
Portage Sites – Example



Source: (Bleakley and Lin, 2012)

return

Figure: Flow Variation in DRC River Network



return

Predicting City Location with Portage Sites

Table: Fractionalization and Portage Probability

	PC	PC	PC	PC
Portage Score	0.107	0.095	0.014	0.008
	[0.010]***	[0.010]***	[0.013]	[0.013]
Dist to River	<100km	<100km	<50km	<50km
River FE	N	Y	N	Y
Mean Dep.	0.397	0.397	0.497	0.497
Observations	36,741	36,741	22,855	22,855

Notes: The fractionalization measures are standardized. The regressions include malaria suitability, land suitability, historic population, ruggedess and river distance as controls, as well as country fixed effects.

return

Predicting City Location with Portage Sites

Table: Portage IV - Predict City

	P(city)	P(city)	P(city)	P(city)
Portage Score	0.035	0.041	0.035	0.043
	[0.003]***	[0.004]***	[0.005]***	[0.005]***
Dist to River	<100km	<100km	<50km	<50km
River FE	N	Y	N	Y
Mean Dep.	0.127	0.127	0.136	0.136
Observations	37,304	37,304	23,000	23,000

Notes: Controls include land suitability, malaria suitability. All regressions include country fixed effects. Fractionalization measures are standardized, and defined using a 50km buffer from the grid centroid. The "Dist" row describes the sample cutoff of distance to nearest river for that particular regression.

return

Predicting City Location with Portage Sites

Table: Portage IV - Prob. Conflict

	P(conflict)	P(conflict)	P(conflict)	P(conflict)
City*PC Fract	0.055	0.063	0.071	0.049
	[0.011]***	[0.011]***	[0.019]***	[0.015]***
PC Fract	-0.007	-0.008	-0.010	-0.006
	[0.002]***	[0.002]***	[0.004]**	[0.003]**
River FE	N	Y	N	Y
Dist to River	<100km	<100km	<50km	<50km
F-stat	31	52	13	31
Mean Dep. Var	0.014	0.014	0.014	0.014
Observations	36,747	36,747	22,861	22,861

Notes: Controls include land suitability, malaria suitability. All regressions include country fixed effects. Fractionalization measures are standardized, and defined using a 50km buffer from the grid centroid. Prob. conflict is defined as the proportion of years in which the grid experienced a conflict across 1975-2021.

return

Table: Portage IV - DHS Durables

	Durables	Durables	Durables	Durables
City*PC Fract	-0.695	1.527	-0.220	-0.541
	[0.080]***	[0.653]**	[0.042]***	[0.091]***
PC Fract	0.406	-1.080	0.089	0.317
	[0.054]***	[0.439]**	[0.029]***	[0.063]***
River FE	N	Y	N	Y
Dist to River	<100km	<100km	<50km	<50km
F-stat	114	7	581	135
Mean Dep. Var	-0.050	-0.050	-0.008	-0.008
Observations	472,821	472,821	308,488	308,488

Notes: Controls include land suitability, malaria suitability. All regressions include country fixed effects. Fractionalization measures are standardized, and defined using a 50km buffer from the grid centroid. Durables is a principle component analysis of assets reported in DHS samples.

return

Table: Portage Score and Hydrological Features

	Portage Score	Portage Score	Portage Score	Portage Score
Discharge Variation	0.021		0.021	
	[0.003]***		[0.003]***	
Flow Variation		0.252		0.442
		[0.068]***		[0.064]***
Dist to River	<50km	<50km	<100km	<100km
Mean Dep. Var	1	1	0	0
Observations	23,219	23,219	37,460	37,460

return